

EXHIBIT I



SARAH R. LAFRENIERE 888 16th Street, N.W.
Associate Suite 300
Washington, DC 20006

(202) 540-7159 Direct
202-540-7200 Main
202-540-7201 Fax

slafreniere@hausfeld.com

March 18, 2021

All Domestic Defendants' Counsel

VIA EMAIL

RE: Defendants' methodologies for technology-assisted review ("TAR") - *In re Diisocyanates Antitrust Litigation*, MDL No. 2862

Dear Counsel:

Plaintiffs write in response to Alden Atkin's March 16, 2021, letter regarding Defendants' TAR methodologies.

I. Richness/Estimation Sample

Your letter asked if the parties remain at issue with respect to Defendants' proposed richness/estimation sample. Plaintiffs do not believe a richness or estimation sample is necessary, but instead request that Defendants adopt Plaintiffs' proposed validation proposal, outlined in Section VII and Appendix A of Plaintiffs' proposed TAR methodologies of February 19, 2021. Plaintiffs have not proposed to include any richness sample in their own methodologies, as our February 19 disclosures convey. As we explained in our March 4, 2021 letter and during our meet and confers of March 1 and 12, 2021, our concerns rest with the entirety of Defendants' validation procedures, and a richness sample can serve "only as a general guidepost."¹ Defendants' base the totality of their review on their imprecise richness sample, and we believe this is statistically improper and inadequate.

Our March 4 letter demonstrated how a margin of error of $\pm 5\%$ can result in absurd and even burdensome results in stopping and validating TAR. While we appreciate Defendants' offer toward compromise, a $\pm 2\%$ margin of error does not ameliorate these concerns. Namely, with the size of the collections the Parties are discussing, a $\pm 2\%$ margin of error reflects, in a collection of one million documents, 40,000 documents ($\pm 2\%$, is 4% of the population, or 40,000). This is too imprecise to be meaningful. Defendants then analyze whether their target

¹ To respond to your contention that Plaintiffs only raised "for the first time" our concerns with the richness sample during the March 12 meet and confer, we note that our own methodologies propose no richness sample, and in our March 4 letter discussing the margin of error for the richness sample we noted "in any event, this is a meaningless argument because Defendants' entire process is fundamentally flawed as described herein."



recall has been achieved based on this imprecise metric. Our March 4, 2020 letter also reflected how Defendants cannot accurately calculate recall on the basis of the richness sample. Namely, if Defendants only collect documents up to the “target” recall based on their prevalence (a/k/a richness) estimate, but the real prevalence is at the upper limit of the margin of error, recall will not be the target 70%, but instead somewhere near 40%. Even with a $\pm 2\%$ margin of error, the recall estimates can be off by a substantial margin.

To demonstrate, suppose the collection contains one million documents and 200,000 are *actually* relevant. The prevalence of this collection would be 20%. If the margin of error were $\pm 2\%$, we could state with confidence only that the prevalence *estimate* was likely to be between 18% and 22%; in other words, that the estimated number of relevant documents was somewhere between 180,000 and 220,000 documents. Suppose the richness sample reports a prevalence estimate of 180,000 (18%) and TAR was terminated when 70% of the estimated number of relevant documents—or 126,000 relevant documents—had been found. The actual recall that would be achieved would be 63% ($126,000/200,000$), not 70% as claimed. Suppose that, equally likely, the estimate was 220,000, and TAR was terminated at 70% of the estimate—when 154,000 documents had been found. The actual recall would be 77%, not 70%. Either way, the recall estimation is inaccurate—and this is the basis for Defendants’ stopping and validating the review.

If Defendants feel it is necessary for their own purposes to initially estimate prevalence, Plaintiffs do not object to that. We simply object to stopping and validating the review on the basis of this measure.

II. Defendants’ Stopping Criteria

Our March 4, 2021 letter explained why Defendants’ stopping criterion is inadequate. We stand by those reasons. We assume that Defendants’ agreement to reduce the margin of error to $\pm 2\%$ also refers to the elusion sampling, and we appreciate that offer toward compromise as well. As noted above, stopping at an “acceptable level of recall” as Defendants propose is inadequate because Defendants cannot determine if they have achieved an acceptable level of recall using the methodologies you have propounded. The methodologies are flawed.

You contend that when Defendants reach an “acceptable level of recall,” TAR 2.0 will prioritize for review those documents most likely to be responsive, and therefore, there is an “infinitesimal” possibility that once achieving the target recall the next tranche of documents will be 90% responsive. We agree with Defendants’ reasoning that it is extremely unlikely that the review will achieve high recall and continue to yield a high proportion of responsive documents. In other words, Defendants should have no problem meeting Plaintiff’s criterion if high recall is in fact achieved, and Plaintiffs should have no problem assuming (pending final validation) that high recall was achieved if the criterion is met.



Using Defendants' estimate of recall (rather than true recall), the possibility of missing a substantial number of relevant documents is not infinitesimal at all, and academic literature has demonstrated that estimates of achieving target recall are often woefully inaccurate.² Indeed, the possibility falls squarely within Defendants' proposed margin of error. Even assuming a ±2% margin of error and the metrics in the above example, if Defendants collected 126,000 documents, when in fact 200,000 responsive documents existed (a measure that fits within the margin of error), it is *highly likely* that the next tranche of documents TAR selects for review are mostly (even 90%) responsive—particularly because Defendants are using TAR 2.0. But there is no need to speak in hypotheticals. The TAR tool will itself provide you with an estimated prevalence for the next tranche of documents. Defendants appear to be stating that they will stop the review when target recall is achieved *even if* the next tranche of documents identified by TAR is largely responsive. If that is the case, how do Defendants intend to approach the scenario where the target recall is purportedly achieved but TAR is still identifying numerous documents as likely to be responsive?

Plaintiffs' proposed stopping criteria is objective, reasonable, poses no additional burden, and takes into account the inability to adequately predict prevalence and recall from the outset. TAR should conclude when the last few batches of documents identified by TAR for human review contain no more than five to ten percent responsive documents, and none of the responsive documents is novel and/or more than marginally relevant. Defendants will produce those last-reviewed responsive documents to Plaintiffs (as they must), and we have confidence in our collective ability to negotiate in good faith about novelty and relevance.³ We ask Defendants to explain how they justify concluding the review if Plaintiffs' proposed criteria are not achieved, in other words, if the last tranche of documents contains more than 10% responsive documents and/or the documents are novel or more than marginally relevant.

III. Plaintiffs' Proposed Validation and Measure of Recall

As we discussed, Plaintiffs believe a robust post-review validation process, such as that proposed by Plaintiffs, is necessary to evaluate the completeness of the review. To reduce

² See, e.g. Maura R. Grossman and Gordon V. Cormack, *Comments on “The Implications of Rule 26(g) on the Use of Technology-Assisted Review”* 7 Fed. Cts. L. Rev. 286, 288, 302

(critiquing TAR validation like Defendants' that focus on target recall as inappropriate when continuous active learning is employed and noting recall “is difficult to measure properly.”)

³ We must however respond to Defendants' contention that the TAR negotiations have been protracted. Any delay rests with Defendants. Plaintiffs asked Defendants if they were using TAR no later than December 1, 2020, Defendants did not answer for two months—until February 5, 2021. Plaintiffs provided a response to Defendants' proposed TAR methodologies one week after Defendants' February 19, 2021 disclosures; it has taken almost three weeks for Defendants to respond to that letter. See Letter from A. Atkins to S. LaFreniere, March 16, 2021.



Defendants' burden, we are willing to compromise in eliminating the richness sample and elusion test from the review altogether only if Plaintiffs' stopping criteria and validation proposals are adopted. Numerically, this would result in sampling of almost the same number of documents (approximately 4,800 in Defendants' proposal and 5,000 in Plaintiffs'). Please let us know if you will agree to this proposal.

Defendants contend that Plaintiffs' calculation of recall "does not provide a statistically meaningful measurement of anything." You cite to the definition of recall as "the fraction of relevant documents that are identified as relevant by a search or review effort." Maura R. Grossman and Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review*, with Foreword by John M. Facciola, U.S. Magistrate Judge, 2013 Fed. Cts. L. Rev. 7 (January 2013), at 27. We agree with this definition, and it accurately states what Plaintiffs' Appendix A measures, as the Appendix was developed from the Broiler Chickens Case, where Maura Grossman served as the Special Master.⁴ Plaintiffs' proposal simply breaks down the document collection into four subcollections, to determine if there is one particular category that is more prone to error. Such *stratified sampling*⁵ is a well-established statistical method to improve accuracy. Specifically, the validation methodology requires a subject matter expert to blindly re-review samples of documents that were (a) coded responsive; (b) deemed by a human to be nonresponsive; (c) deemed by TAR to be nonresponsive; and (d) excluded from TAR and not reviewed. The number of responsive documents in the last three categories represent a statistically valid estimate of the number of responsive documents not identified. Then, recall is estimated by comparing the number of documents found to a statistically valid estimate of the number of documents not identified.

Plaintiffs' measure of recall reflects what courts and scholars alike understand to be recall: the percentage of responsive documents identified within the document collection—*i.e.* the whole document collection.⁶ We are unaware of any case where recall is understood differently, unless they are in error. If you have other examples, please provide them to us.

We are not sure if your statement that it is "statistically meaningless to perform the sample on the raw collection of documents when the review is being performed on a different set of documents" refers only to the richness/estimation sample, or to Plaintiffs' proposed validation.

⁴ Order Regarding Search Methodology for Electronically Stored Information, *In re: Broiler Chicken Antitrust Litig.*, No. 1:16-cv-8637 (N.D. Ill. Jan. 3, 2018).

⁵ https://en.wikipedia.org/wiki/Stratified_sampling

⁶ Grossman and Cormack, *Comments on "The Implications of Rule 26(g) on the Use of Technology-Assisted Review"* 7 Fed. Cts. L. Rev. at 288 ("validation . . . is best achieved by considering the end-to-end effectiveness of the review, and evaluating the totality of the evidence derived from multiple sources, not by considering only a single target measure applied to a particular phase of the review process.")



As to the former, we have expressed our doubts about the utility of the richness sample, and have no opinion regarding the population from which it is drawn, so long as it is used only to estimate review effort, and plays no role in stopping or validation. As for Plaintiffs' validation, which includes documents not included in the TAR review population, this is both necessary and ordinary. Defendants' obligations under Rule 26(g)(1) to conduct a reasonable inquiry includes determining whether responsive documents were unreasonably culled from the review population (through search terms, TAR, or otherwise). Unless Defendants are representing that there are *no* responsive documents outside the TAR review population, then an accurate validation of Defendants' review includes consideration of all potential sources of missed documents, including the documents not subject to TAR. Courts have long held that sampling of a null set (documents that did not hit on search terms) was appropriate.⁷ Just because TAR is now being applied as a further culling step after search-term culling does not negate the validity of those cases.

Finally, we respond to your contention that the "the procedure invites disputes about the responsiveness determinations made during the Validation Protocol." We are not sure how. Plaintiffs have only asked Defendants to provide (a) the responsive documents identified during the validation (which we are entitled to); (b) a table indicating the results; and (c) a calculation of recall using the formula in Appendix A. It is theoretically possible that if Defendants improperly identify many non-responsive documents as responsive (*i.e.*, false positives), the precision of the review could be called into question—but Defendants have an obligation to ensure their review is sufficiently precise.⁸ Unless the precision is very low, there should be no issue because human review after TAR will increase precision.⁹ However, because Defendants are not providing Plaintiffs with non-responsive documents, Plaintiffs have no insight into whether documents were inappropriately coded as non-responsive (*i.e.*, false negatives). The only insight Plaintiffs glean from the proposed validation is an accurate measure of recall, which, if low, Defendants have an independent obligation to remedy.

⁷ *City of Rockford v. Mallinckrodt ARD Inc.*, 326 F.R.D. 489, 494-95 (N.D. Ill. 2018) (ordering defendants to provide Plaintiffs with a sample of documents from the null set after determining "a random sample of the null set will help validate the document production in this case."); *In re Seroquel Prod. Liab. Litig.*, 244 F.R.D. 650, 622 (M.D. Fla. 2007) ("Common sense dictates that sampling and other quality assurance techniques must be employed to meet requirements of completeness.").

⁸ *In re Domestic Airline Travel Antitrust Litig.*, No. 15-1404, 2018 WL 4441507, at *4-5 (D.D.C. Sept. 13, 2018) (extending deadline for the conclusion of fact discovery when Defendants' production resulted in low precision).

⁹ See Maura Grossman & Gordon Cormack, Gordon, "Navigating Imprecision in relevance Assessments on the Road to Total Recall: Roger and Me," <https://cormack.uwaterloo.ca/roger/roger.pdf>.



PAGE 6

We look forward to discussing further on March 19, 2021.

Very truly yours,

/s/ Sarah R. LaFreniere

Sarah R. LaFreniere

cc: Megan E. Jones
Jason S. Hartley
Jason Lindner
William Pietragallo, II